Genomic Sequence of *Heliothis virescens* Ascovirus 3g Isolated from *Spodoptera exigua*

Guo-Hua Huang, Yun-Sheng Wang, Xing Wang, Tyler A. Garretson, Liang-Ying Dai, Chuan-Xi Zhang, and Xiao-Wen Cheng

Hunan Provincial Key Laboratory for Biology and Control of Plant Diseases and Insect Pests, Hunan Agricultural University, Changsha, Hunan, People’s Republic of China; Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha, Hunan, People’s Republic of China; Department of Microbiology, Miami University, Oxford, Ohio, USA; and Agriculture Key Laboratory of Agricultural Entomology, Institute of Insect Science, Zhejiang University, Hangzhou, Zhejiang, People’s Republic of China

*Heliothis virescens ascovirus 3a* (HvAV-3a), a member of the family *Ascoviridae*, has the highest diversity among ascovirus species that have been reported in Australia, Indonesia, China, and the United States. To understand the diversity and origin of this important ascovirus, the complete genome of the HvAV Indonesia strain (HvAV-3g), isolated from *Spodoptera exigua*, was determined to be 199,721 bp, with a G+C content of 45.9%. Therefore, HvAV-3g has the largest genome among the reported ascovirus genomes to date. There are 194 predicted open reading frames (ORFs) encoding proteins of more than 50 amino acids. In comparison to HvAV-3e reported from Australia, HvAV-3g has all the ORFs in HvAV-3e with 6 additional ORFs unique to HvAV-3g, including 1 peptidase C26 gene with the highest identity to *Drosophila* spp. and 2 gas vesicle protein U (GvpU) genes with identities to *Bacillus megaterium*. The five unique homologous regions (hrs) and 25 baculovirus repeat ORFs (bro) of HvAV-3g are highly variable.

Insect-specific ascoviruses (AVs) in the *Ascoviridae* family have double-stranded circular DNA with a reported genome size range of 100 to 190 kbp (4, 15). Among the four official ascovirus species of the *Ascoviridae* family, including *Spodoptera frugiperda* ascovirus 1a (SfAV-1a), *Trichoplusia ni* ascovirus 2a (TnAV-2a), *Heliothis virescens ascovirus 3a* (HvAV-3a), and *Diadromus pulchellus* ascovirus 4a (DpAV-4a), HvAV-3a has the widest distribution worldwide. HvAV-3a variants include HvAV-3b, HvAV-3c, HvAV-3d, and HvAV-3f, reported in the United States, HvAV-3e in Australia, HvAV-3g in Indonesia, and HvAV-3h in China (2, 7–9, 12), whereas SfAV-1a and TnAV-2a have been reported only in the United States (6, 10).

HvAV-3g was initially named *Spodoptera exigua ascovirus 5a* (SeAV-5a) but was later confirmed by DNA-DNA hybridization studies to be a variant of HvAV-3a (4, 7). The HvAV-3g virions were purified from the hemolymph of *S. exigua* larvae infected with HvAV-3g, and the DNA was isolated from the purified virions by proteinase K digestion and CsCl-ethidium bromide gradient ultracentrifugation (8). The genomic DNA was sheared into fragment sizes of 400 to 600 bp by ultrasonication and sequenced by the Solexa genome analyzer at the Beijing Genome Institute, Shenzhen, China (BGI). We obtained a total of 6,077,778 clean paired-end (PE) reads with an average insert fragment size of 500 bp for approximately 547 million nucleotides, representing a 2,700× coverage of the 183-kbp HvAV-3g genome that was estimated by restriction fragment length polymorphism (RFLP) analysis (8). A total of 300,000 PE reads were used to assemble a single contig by Edena (11). The assembled contig representing the entire HvAV-3g genome sequence was confirmed by comparing six predicted restriction fragment profiles from the genome, for HindIII, EcoRI, BamHI, XbaI, Xhol, and PstI, with the matching fragment profiles produced by actual restriction digestions of the HvAV-3g genome (7, 8, 12).

DNA sequence analysis showed that HvAV-3g has a genome size of 199,721 bp, which is the largest among all the reported ascovirus genomes to date (2, 3, 5, 14). The HvAV-3g genome has a G+C content of 45.9%, which is similar to that of HvAV-3e (45.8%) but dissimilar to that of SfAV-1a (49.2%), TnAV-2c (35.4%) and DpAV-4a (49.6%) (2, 3, 5, 14). There are 194 open reading frames (ORFs) encoding predicted proteins of more than 50 amino acids.

Of its 194 ORFs, 188 of the HvAV-3g ORFs showed homology to that of HvAV-3e (2). The remaining 6 unique ORFs from HvAV-3g include a peptidase C26 gene with the highest identity to *Drosophila* spp. and two gas vesicle protein U (GvpU) genes that are found in the bacterium *Bacillus megaterium*. The peptidase also has Y-glutamyl hydrolase activity in humans (1), and GvpU is required to produce gas vesicles in *B. megaterium* (13). These two genes were not found in the other reported ascoviruses, including the closely related HvAV-3e (12). Five unique homologous regions (hrs) and 25 baculovirus repeated ORFs (bro) which are highly variable, were found in the genome of HvAV-3g.

**Nucleotide sequence accession number.** The GenBank accession number of *Heliothis virescens ascovirus 3g* is JX491653.

**ACKNOWLEDGMENTS**

This work was supported by the National Natural Science Foundation of China (31228020) and the National Major Science and Technology Project of the twelfth Five-year Plan (2012BAD27B00).

**REFERENCES**


Received 28 August 2012 Accepted 28 August 2012

Address correspondence to Xiao-Wen Cheng, chengx@muohio.edu, or Liang-Ying Dai, dail@hunau.net.

Guo-Hua Huang and Yun-Sheng Wang contributed equally to this work.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/M.02342-12